

Hướng dẫn gán nhãn dữ liệu aspect-sentiment cho chủ đề Nhà Hàng (Restaurant)

1. Mục tiêu

Định hướng:

- ❖ Với mỗi đầu vào là review (bản nhận xét đánh giá của khách hàng), mục tiêu là xác định các bộ (Aspect Category, Opinion Polarity).
- ❖ Dữ liệu này được xây dựng tương tự như sub-task2 của task 5
Task 5: Aspect-Based Sentiment Analysis

Mục tiêu cụ thể:

- ❖ Gán nhãn ít nhất 3000 review
- ❖ Dữ liệu thực tế của khách hàng, có thể tham khảo từ: <https://www.foody.vn>

2. Các loại nhãn

A) Nhãn đánh giá, quan điểm, cảm nhận (**opinion polarity**) bao gồm:

positive: tức nhận xét tích cực, khen

negative: tức nhận xét tiêu cực, chê

neutral: tức nằm giữa tích cực và tiêu cực (trung dung)

B) Nhãn **aspect**: gồm bộ **entity-attribute** của thực thể

Các loại **Entity**:

restaurant (nhà hàng): nói đến nhà hàng nói chung, tổng kết
+ khi không chỉ cụ thể một entity nào như ở dưới, thì có nghĩa là đề cập chung đến restaurant

+ hầu hết các review chúng ta đều có thể xác định sentiment cho entity này.

ambience (không gian): không gian của nhà hàng, liên quan đến giải trí, bên trong có thoáng, dễ chịu không? Ví dụ “nhà hàng bố trí đẹp, có sân vườn rộng, có chỗ cho trẻ em chơi”

location (vị trí): liên quan vị trí nhà hàng có thuận tiện không, view nhà hàng có ok không, có chỗ gửi xe không, ..

food (đồ ăn): đồ ăn, các món ăn

service (phục vụ): người phục vụ, cách phục vụ khách hàng

drinks (đồ uống): đồ uống

Các loại **Attribute**:

general (nói chung): thường dùng với restaurant

quality (chất lượng): thường dùng với food, drink

price (giá cả): giá của restaurant, hoặc của food, drink, service,

style_option (kiểu, tùy chọn): dùng cho cách trình bày; kiểu phục vụ;

lựa chọn trong thực đơn có phong phú không;

miscellaneous (thuộc tính khác): không thuộc 4 cái trên

Chúng ta thường thấy một số aspect như sau:

	GENERAL	PRICES	QUALITY	STYLE & OPTIONS	MISCELLANEOUS
RESTAURANT	✓	✓	x	x	✓
FOOD	x	✓	✓	✓	x
DRINKS	x	✓	✓	✓	x
AMBIENCE	✓	x	x	x	x
SERVICE	✓	x	x	x	x
LOCATION	✓	x	x	x	x

restaurant#general
restaurant#price
restaurant#miscellaneous

food#quality
food#stype_option
food#price

drinks#quality
drinks#stype_option
drinks#price

service#general
embience#general
location#general

3. Cách gán nhãn

Cách 1:

#id

Text

<aspect 1> positive hoặc negative hoặc neutral

<aspect 2> positive hoặc negative hoặc neutral

...

#id

...

Cách 2:

Dạng xml

=====

Ví dụ:

#1

180K/suất không phải là cái giá rẻ đối với người Sài Gòn. Nhưng nếu so với giá Chả cá Anh Vũ - Giảng Võ và vị trí ngay khu TT Q3 thì cũng hợp lý. Phải nói là chả có ngon nhất SG mình từng ăn

tới lúc này. Bài trí và hương vị khá giống nguyên bản ngoài Hà Nội. Từ chanh ớt, vị mắm tôm cho tới cái chảo rán cá. Duy chỉ có mùi ngũ vị hương hơi đậm là mình ko khoái so với ngoài kia. Creme Brulee khá ngon và ít ngọt. Nói chung sẽ quay lại nhiều.

<restaurant#price> neutral

<food#quality> positive

<food#style_option> positive

<restaurant#general> positive

#2

Hôm nay mình với 3 bạn nữa tới ăn. Quán thấy khá đông, vị trí ngay trung tâm, không gian bình thường. Máy bạn nhân viên thấy khá lúng túng trong việc phục vụ, phải nhắc 2 3 lần mới được. Ăn thì thấy cũng được thôi, có lẽ do mình không thích món Bắc lẩu. Quán thì chỉ có đúng 1 món thôi, mình kêu 4 phần ra được 2 chảo cá không nhiều lắm. Bánh trứng thì ngon. 4 bạn ăn gần hết 900k. Giá quá mắc cho món này. Do mọi người khen nhiều nên ăn 1 lần cho biết thôi.

<food#price> negative

<service#general> negative

<ambience#general> neutral

<location#general> positive

<food#quality> positive

<restaurant#general> negative

4. Tài liệu tham khảo

<http://alt.qcri.org/semeval2016/task5/>

Task 5: Aspect-Based Sentiment Analysis

Subtask 2: Text-level ABSA. Given a set of customer reviews about a target entity (e.g. a laptop or a restaurant), the goal is to identify a set of {aspect, polarity} tuples that summarize the opinions expressed in each review.

GÁN NHÂN CỤ THỂ

(Vietlex Team)

TẦNG 1: ENTITY

1. Restaurant:

- Nói đến nhà hàng nói chung. Vd:

Quán mở từ 5h sáng tới tầm 10h30 là bán gần hết sạch.

- View nhà hàng

- Khen, chê, đánh giá mà không có chủ ngữ (mà không nói về vị trí hay con người). Vd:

Nói chung là không đạt ...

- Những câu có dạng:

xxxx

2. Food:

- Tất cả đồ ăn/món ăn trong nhà hàng (kể cả kem, chè, hoa quả)

- Cách thức trình bày, phục vụ, chất lượng, lựa chọn...

3. Drinks:

- Tất cả đồ uống trong nhà hàng,...

- Cách thức trình bày, phục vụ, chất lượng, lựa chọn...

4. Ambience:

- Nói đến không gian chung của nhà hàng, như: thiết kế bên trong như thế nào, khu vực giải trí ra sao. Vd:

nhà hàng bố trí đẹp, có sân vườn rộng, có chỗ cho trẻ em chơi.

5. Service:

- Kỹ năng, thái độ phục vụ

- Những từ gắn với yếu tố con người (thân thiện, vui vẻ, lịch sự...)

6. Location:

- Vị trí, địa điểm của nhà hàng có thuận tiện không

- Chỗ gửi xe

TẦNG 2: ATTRIBUTE

1. Nhân GENERAL (áp dụng cho các Entity 1, 4, 5, 6)

Entity	Attribute
1. Restaurant 4. Ambience 5. Service 6. Location	General

1. Restaurant#General

- Câu không có thực thể rõ ràng. Vd:

Còn lại tất cả đều tốt.

Nói chung là ưng

- Câu nhận xét chung về nhà hàng (trừ vị trí địa điểm, con người). Vd:

Quán khá đông

Quán này khá nổi tiếng

Quán này lâu năm lắm rồi í

Quán này mình hay ăn cực

Nchug mình sẽ ko quay lại

khuyến các bạn nên đến

Ở đây khá là đông

rất ổn để ghé lại

Popeyes là quán gà rán mình thích nhất phù hợp cho hội nhóm gặp mặt bạn bè

4. Ambience#General

- Thực thể là khu vực trong khuôn viên nhà hàng, bao gồm sân, vườn, khu vui chơi, v.v. Vd:

trong phòng máy lạnh

quán ngồi siêu mát mẻ luôn

Quán khá nhỏ nhưng nhìn rất sang và ấm cúng

5. Sevice#General. Vd:

Phục vụ tốt/không tốt

Nhân viên thân thiện, lịch sự, dễ thương, cau có, khó gần, chuyên nghiệp

Thái độ phục vụ

Kịp thời, nhanh chóng...

Bàn ghế hơi dầu mỡ

6. Location#General:

Nhà hàng có địa điểm, vị trí...: ngay ngoài ngõ

Nhà hàng gần/xa trung tâm...

2. Nhãn PRICES (áp dụng cho các Entity 1, 2, 3)

Entity	Attribute
1. Restaurant	Prices
2. Food	
3. Drinks	

1. Restaurant# Prices

- Giá của nhà hàng, thực thể là nhà hàng. Vd:

Nhà hàng có giá rẻ

Nhà hàng có chất lượng (không) phù hợp/(không) tương xứng giá tiền

Nhà hàng là 1 sự lựa chọn tiết kiệm

Giá Bình Dân

Giá rổ okay

Điểm tụ tập giá rẻ

Giá 17k, quán này giá rẻ, đa số đều 17k hết

Giá ở đây tuy hơi cao tầm 40-45k
 Địa điểm để ăn đồ Nhật giá rẻ chính là ở đây rồi nè
 Quán bình dân nhưng giá nhà hàng
 Nhìn từng này nhưng được giảm 70% nên ăn chả hết bao nhiêu
 Đang có chương trình khuyến mại giảm 50%
 Đc cái quán có nhiều chương trình khuyến mãi
 tặng 1 thẻ tích điểm
 Hay có các chương trình khuyến mãi

- Giá tính chung cả cho đồ ăn và đồ uống. Vd:
 tầm chục người no nê có gọi đồ uống hết tầm 1.900.000 đ
 2 ng đi ăn thêm cả tráng miệng khoảng 100k là oke rồi
 Trung bình 2 bạn khoảng 250k
 Chỉ với 30k bạn sẽ có một bát miến trộn siêu nhiều và 1 cốc sữa đậu cực ngon

2. Food# Prices

- Giá món ăn, bữa ăn. Vd:
 xxx

3. Drinks# Prices:

- Giá các đồ uống. Vd:
 xxx
 xxx

3. Nhãn QUALITY (áp dụng cho các Entity 2, 3)

Entity	Attribute
2. Food 3. Drinks	Quality

2. Food#Quality

- Thực thể là món ăn, bữa ăn
 đẳng cấp, sang trọng
 chất lượng tốt/tệ
 đạt/không đạt, xứng đáng/không xứng đáng
 ngon/ dở, ổn, được, cũng ổn

3. Drinks#Quality:

- Thực thể là đồ uống. Vd:
 đẳng cấp, sang trọng
 chất lượng tốt/tệ
 đạt/không đạt, xứng đáng/không xứng đáng

4. Nhãn STYLE&OPTIONS (áp dụng cho các Entity 2, 3)

Entity	Attribute
2. Food 3. Drinks	Style&Options

2. **Food#Style&Options.** Vd:

Cách bày biện, chế biến đồ ăn: đúng chuẩn bún bò Huế, độc đáo

Nhiều/ít lượng đồ ăn: đồ ăn đầy đặn, Lèo tèo vài sợi là hết, Miếng gà to, 1 bàn tiệc hoành tráng, Siêu nhiều thịt bò, hải sản ăn kèm khá ít

Nhiều/ít món, lựa chọn: Menu khá phong phú, Có nhiều loại vị khác nhau

3. **Drinks#Style&Options.** Vd:

Cách bày biện, chế biến đồ uống

Nhiều/ít món, lựa chọn

Nhiều/ít lượng đồ uống: nước thì được uống thêm thoải mái

5. Nhãn MISCELLANEOUS (chỉ áp dụng cho Entity 1)

1. **Restaurant#Miscellaneous.** Vd:

Nhà hàng an toàn/không an toàn

Nhà hàng an ninh đảm bảo

Nhà hàng có wifi

Giờ đóng/mở cửa

hoá đơn

đường vào

hợp vệ sinh

Không thích là tính tiền giữ xe 3k

Ko có ai coi xe

Giữ xe miễn phí

nhân được mặc đồng phục theo phong cách áo thi đấu

TẦNG 3: VALUE (áp dụng cho tất cả 6 Entity)

1. **Restaurant:**

- positive: *tốt, đẹp, được, sạch sẽ, ok, thoáng, mát, tuyệt vời,...*

- negative: *xấu, tồi, tệ, bẩn, chưa được tốt, bí, ẩm thấp, không đẹp, không đầy đủ, cũ, không sang trọng...*

- neutral: *tạm được, bình thường, không có gì đặc sắc,...* + có 2 ý được đánh giá trái ngược positive & negative

2. **Location:**

- positive: *trung tâm, gần, tốt, đẹp, ok, dễ tìm, dễ thấy, ngay...xxx, hoặc ở một khu vực trung tâm, tiện lợi: “gần nhà”, “gần chỗ làm” “ở khu phố cổ”, “nằm trong khu vực ẩm thực”, “nằm ngay khu phố tây”, “khu...sâm uất”; “Tính ra ăn 4 người no ứ hự ở cái quán trên q1 vậy là quá rẻ”; “Ở quận 1 nhưng giá cũng ok lắm”*

- negative: *xa, khó tìm, khó thấy*

- neutral: *tạm được, bình thường + có 2 ý được đánh giá trái ngược positive & negative, hoặc , hoặc 1 địa chỉ cụ thể (mà không có đánh giá kèm theo)*

3. **Service:**

- positive: *tốt, được, thân thiện, vui vẻ, lịch sự, chuyên nghiệp, ok*

- negative: *kém, tồi, không chuyên nghiệp, thiếu chuyên nghiệp, không vui vẻ, hơi chậm,*

- neutral: *tạm được, bình thường, đông người nên đợi hơi lâu + có 2 ý được đánh giá trái ngược positive & negative*

4. Food&Drinks

- positive: ngon, tuyệt, hợp khẩu vị, tốt, được, tương đối, đáng tiền, hấp dẫn, ok, lạ
- negative: dở, không ngon, không hợp khẩu vị, mặn/nhạt, hơi ít
- neutral: chấp nhận được, cũng ổn, cũng được, ngon nhưng không bỏ, tạm ổn, tạm được, bình thường + có 2 ý được đánh giá trái ngược positive & negative

5. Price

- positive: rẻ, tốt, được, ổn, hợp lí, rất hợp lí, phải chăng, phù hợp, đáng tiền, khá mềm, vừa phải, hấp dẫn, ok, miễn phí, free, ko hề mắc chút nào, tương xứng, có/chi/chi có/mà có/mà chi xxx, có/chi xxx thôi, 2 người chỉ 70k là quá ok luôn
- negative: đắt, quá đắt, cao, quá cao, hơi cao
- neutral: chấp nhận được, , cũng ổn, cũng được, tạm ổn, tạm được, không quá đắt, vừa phải, giá vừa tầm, bình thường, hơi đắt chút mà xứng đáng vì đồ ăn cho bao la, hơi cao nhưng chất lượng, có 2 ý được đánh giá trái ngược positive & negative, chỉ nêu giá tiền cụ thể chứ không có đánh giá

Training Data

Reviews không làm:

37#bánh gạo cay #take away #yumy :#3

57#oursaigon #bachtuotnuong #streetfood lần đầu tiên thấy or quán bình dân con bạch tuộc to vậy.

84•Hot spicy chicken ...

106#□□ Dầu dừa ...

107#□□ Ring Ring Ring*** chuẩn bị cho máy tóc dài ăn tết đi nào các nàng ơi ***□ Sản phẩm làm đẹp Tóc: TINH DẦU BUỒI..

488#Bị cuồng □□□□□□□

502#Black, Hand and Sobi craker #lozi #HAND

1051#friendly - delicious tasty

2667#Sống trên đời không ăn miếng dồi chó. Chết xuống âm phủ biết có hay không

2995#Yum □

2996#Yummy

2997#yummy 29/10 □□□

Development Data

Reviews không làm:

377#Ai cũng thích nộm bò khô, ai cũng thích được ăn cái món trộn này vào buổi chiều, và công bằng mà nói thì món nộm bò khô chỉ nên ăn từ xế chiều trở đi. Bạn không thể ăn nó vào buổi sáng thay bún phở, cũng chẳng nên ăn nó vào buổi trưa bởi nó chẳng... giải quyết được gì cho cơn đói của bạn. Vậy nên thời điểm đẹp nhất để ăn nộm là buổi chiều, và đúng vào khoảng thời gian xế chiều đó, ăn nộm mới thấy ngon hơn cả.

1157@Cách làm bún chả đơn giản, nhưng để làm được bát bún chả ngon còn phụ thuộc nhiều vào kinh nghiệm và bí quyết gia truyền, đặc biệt quan trọng là cách pha nước chấm của mỗi người. Điều này giải thích vì sao ăn bún chả Hà Nội sẽ khác hẳn so với ăn món này ở nhiều nơi khác.

* Lưỡng lự:

283: Đi cuối tuần quán đông nghẹt, có khi còn phải đứng chờ

300: quán lúc nào cũng đông

2155: Mỗi tội điêm trừ là đũa bần = {RESTAURANT#MISCELLANEOUS,

- View KS nếu để Location thì view của phòng không biết để vào đâu?
- Nếu để View là General thì phần Nhà hàng cũng phải chuyển từ Location sang General?
- Phần Nhà hàng một số thông tin đề nghị xếp như sau:
 - + Quán/Nhà hàng sạch sẽ = {AMBIENCE#GENERAL,
 - + đồ ăn sạch sẽ, vệ sinh = {FOOD#QUALITY,
 - + Không có ai coi xe / gửi xe miễn phí / gửi xe mất 5k... khăn ướt 2k... = {RESTAURANT#MISCELLANEOUS,
 - + giá đồ ăn/ đồ uống chỉ nêu số tiền cụ thể, không đánh giá = {FOOD#PRICES, neutral}
 - + địa chỉ nêu cụ thể, không đánh giá:
 - nếu ngoài đường thì positive}
 - nếu trong ngõ, hẻm thì neutral}